

Citation for published version:

Hoque Tania, M, Kaiser, MS, Abu-Hassan, K & Hossain, MA 2020, 'Pathological test type and chemical detection using deep neural networks: a case study using ELISA and LFA assays', *Journal of Enterprise Information Management* . <https://doi.org/10.1108/JEIM-01-2020-0038>

DOI:

[10.1108/JEIM-01-2020-0038](https://doi.org/10.1108/JEIM-01-2020-0038)

Publication date:

2020

Document Version

Peer reviewed version

[Link to publication](https://doi.org/10.1108/JEIM-01-2020-0038)

The final publication is available at Emerald via <https://doi.org/10.1108/JEIM-01-2020-0038>

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Pathological Test Type and Chemical Detection Using Deep Neural Networks: A Case Study Using ELISA and LFA Assays

Marzia Hoque Tania
*Institute of Biomedical Engineering
Department of Engineering Science
University of Oxford
Oxford, United Kingdom
ORCID: 0000-0002-4496-1896*

M. Shamim Kaiser
*Institute of Information Technology
Jahangirnagar University
Savar, Bangladesh
ORCID: 0000-0002-4604-5461*

Kamal J. Abu-Hassan
*Department of Physics
University of Bath
Bath, United Kingdom
kjaah20@bath.ac.uk
ORCID: 0000-0002-1982-1306*

M A Hossain
*School of Computing and Digital Technologies
Teesside University
Middlesbrough, United Kingdom
a.hossain@tees.ac.uk*

Abstract—

Purpose

The gradual increase in geriatric issues and global imbalance of the ratio between patients and healthcare professionals has created a demand for intelligent systems with the least error-prone diagnosis results to be used by less medically trained persons and save clinical time. This paper aims at investigating the development of an image-based colourimetric analysis. The purpose of recognising such tests is to support wider users to begin a colourimetric test to be used at homecare settings, telepathology, etc.

Design/methodology/approach

The concept of an automatic colourimetric assay detection is delivered by utilising two cases. Training Deep Learning (DL) models on thousands of images of these tests using transfer learning, this paper i) classifies the type of the assay, and ii) classifies the colourimetric results.

Findings

This paper demonstrated that the assay type can be recognised using DL techniques with 100% accuracy within a fraction of a second. Some of the advantages of the pre-trained model over the calibration-based approach are robustness, readiness and suitability to deploy for similar applications within a shorter period of time.

Originality/value

To the best of our knowledge, this is the first attempt to provide Colourimetric Assay Type Classification (CATC) using DL. Humans are capable to learn thousands of visual classifications in their life. Object recognition may be a trivial task for humans, due to photometric and geometric

variabilities along with the high degree of intra-class variabilities it can be a challenging task for machines. However, transforming visual knowledge into machines, as proposed, can support non-experts to better manage their health and reduce some of the burdens on experts.

Plain Language Summary for Kudos

The colourimetric method can tell you valuable information about someone's health status, quality of water, presence of chemical substances etc. by chemical and biochemical reactions. Bringing these frequently used techniques on to an intelligent and digital platform where a user can just take a photograph of the test using his/her mobile phone and attain a faster automatic result can assist the non-experts to manage their health and save time. This research demonstrated such a system using advance algorithms that produce a highly accurate real-time result on the mobile platform by using two examples- 1) a deadly bacterium called tuberculosis (TB) and 2) a pH, i.e. potential of Hydrogen test that can tell if a biological or chemical sample is acidic or alkaline.

Keywords

Computer Vision; Machine Learning; Colourimetric Test; Pre-trained Model; Point-of-Care System; Diagnosis

1. Introduction

Deep Learning (DL) based pathological image analysis is employed in detecting diseases from cold to cancer (Arvaniti *et al.*, 2018; Mane, Viraj; Yakub, 2018). Among many DL methods, the high accuracy of Convolutional Neural Network (CNN) based models trained over millions of images (ImageNet, 2016) has opened the door to utilise transfer learning, expressed through pre-trained models (e.g., AlexNet (Krizhevsky, Sutskever and Hinton, 2012), GoogLeNet (Szegedy *et al.*, 2015), Inception (Szegedy *et al.*, 2016) and ResNet (He *et al.*, 2016)) for a number of applications such as retinopathy (Sahlsten *et al.*, 2019), musculoskeletal disease (Tiulpin *et al.*, 2019), materials properties (Nash, Drummond and Birbilis, 2018) etc. CNNs also find its usefulness in classifying and detecting pathological images (Wang *et al.*, 2018). However, very few research works have been reported with colour based focus on healthcare applications using pre-trained models, which is the main aim of this paper.

Table 1: Application of colourimetric tests across disciplines

Field	Application	Reference
Drug discovery	Structural relationship between materials	Facchini et al. (2018)
Environmental monitoring and quality control	Mercury in water	Chen et al. (2016), Wei et al. (2014)
	Indoor air quality	Qin et al. (2015)
	Iron test in soil and water	Choodum, Sriprom and Wongniramaikul (2019)
	Pesticides in water	Sicard et al. (2015)
	pH and nitrate in water	Lopez-Ruiz et al. (2014)
	Water hardness	Bhattacharjee, Jiang and Behdad (2015)
Food industry	Antibiotics abuse in animals	Yan et al. (2018)
	Bacteria in food	Zheng et al. (2018)
	Food allergen	Coskun, Ahmet F. Wong et al. (2012)
	Iron detection in white wine	Santos Neto et al. (2018)
	Meat quality	Magiati et al. (2019)
	Pasteurisation of milk	Yu et al. (2015)
	Age of bloodstain	Shin et al. (2017)

Forensic science and criminal investigation	Cocaine detection	Smith et al. (2014), Cooper et al. (2012)
	Driving after alcohol intake	Kim et al. (2017)
Materials science	Arc-welding	Serrano et al. (2016)
Medical diagnosis and healthcare	Dengue	Matthews et al. (2012)
	Diabetes	Arnett et al. (2016), Wang <i>et al.</i> (2020)
	HIV	de la Rica and Stevens (2012)
	Kidney diseases	Akraa et al. (2018)
	Malaria	Wongsrichanalai et al. (2007)
	Pregnancy and complications	Bu et al. (2018), Konnaiyan et al. (2017)
	Prostate Cancer	Koo, Wee and Trau (2016), Welch and Albertsen (2009), de la Rica and Stevens (2012)
	Urinalysis	Smith et al. (2016)
Textile	Smart-textile	Promphet et al. (2019)

Colourimetric assays show quantifiable colour change using reagents and chemical constituent. These assays are utilised across a wide range of disciplines as exemplified in Table 1. Depending on the application, purpose and method, assay type can be of many forms. More variations can cause due to the manufacturer of the assay. Such variations, whether visually distinctive or not, can be challenging for less medically trained personnel. Various traditional machine learning approaches are utilised in colourimetric detection (Table 2). However, no DL-based approach, such as CNN, has been suggested for performing any colourimetric test.

Table 2: Colourimetric classification algorithms and their performances

Reference	Classes	ΔE	ML	Performance metrics	Accuracy (%)	Execution time (s)
Alankus et al. (2018)	6x4	Yes	N/a	Confusion matrix; detection accuracy; precision; recall; F1-score	76-100	Rapid
Solmaz et al. (2018)	2; 6; 11	No	LS-SVM, RF	Cross validated accuracy, success rate	90.3; 95	
Rahmat et al. (2018)	10*	Yes	N/a	Accuracy	95.45	N/m
Kim et al. (2017a)	5,9	No	LDA, SVM, ANN	Cross validated accuracy, PPV, NPV	80-100	~9.3
Mutlu et al. (2017)	15	No	LS-SVM	Accuracy; ROC curve	100	N/m
Wang et al. (2016)	13x6	Yes	LDA, PLS	Cross-validated accuracy	100	Rapid
Feng et al. (2014b)	2	No	SVM	k-1 cross validated accuracy, correct vs incorrect	100	8

ΔE : Colour difference; ML: Machine learning; N/a: Not applicable; N/m: Not mentioned; LS-SVM: Least-squares Support-Vector Machines; RF: Random Forest; LDA: Linear Discriminant Analysis; SVM: Support-Vector Machines; ANN: Artificial Neural Networks; PLS: Partial Least Squares; PPV: Positive predictive value; NPV: Negative predictive value; ROC: Receiver operating characteristics s: second

*Each of the classes can be further categorised into different concentration levels or subclasses.

To the best of our knowledge, Tania et al. (2019) have attempted to define colourimetric assay types from the eyes of a machine (i.e. computer vision) and perform any colourimetric test using CNN-based DL for the first time. In this work, pre-trained models have been chosen to determine the type of colourimetric pathological and chemical tests (Tania et al., 2019). The compatibility issue of the proposed detection method with the ASSURED (accuracy, specificity, sensitivity, user-friendliness, robust, rapid, deliverable) criteria is also taken into account. More details on the image-based ASSURED test are also provided in (Hoque Tania et al., 2020).

The objective of this paper is to provide the DL-enabled pathological test type detection using the example of a wet-chemical based Enzyme-Linked ImmunoSorbent Assay or ELISA (Shabut et al., 2018) and a dry chemical based Lateral Flow Assay or LFA (Hoque Tania et al., 2020). Followed by the binary classification of the assay type, this paper extends the experiment to provide the colourimetric classification of the assay. This phenomenon can be also considered as the intra-class classification. For the proof-of-concept, this paper utilises the intra-classes of the LFA used for the assay type detection and classifies it to eight sub-class labels.

Drawing from the literature, at first, this paper provides an outline of the colourimetric pathological and chemical test types in Section 2. In Section 3, the research design and methods are explained. Section 4 presents the findings of this paper, along with a critical analysis, which is then summarised in Section 5.

2. Literature Review and Research Synthesis

An assay is an investigative procedure, especially in laboratory medicine and biochemistry, for qualitatively assessing or quantitatively measuring the presence, amount, or functional activity of a target entity (World Health Organization, 2009; Chang *et al.*, 2019). Assays can be used to determine the components of a substance or object (Wen *et al.*, 2019; Pena-Pereira *et al.*, 2020).

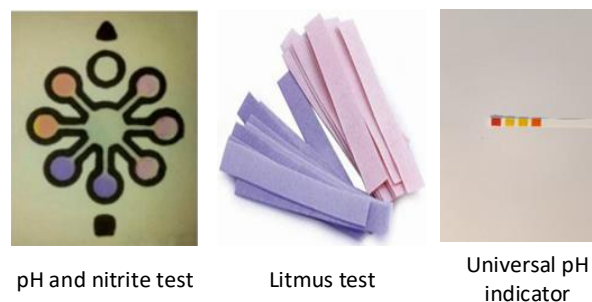


Figure 1: Paper-based pH tests using different approaches

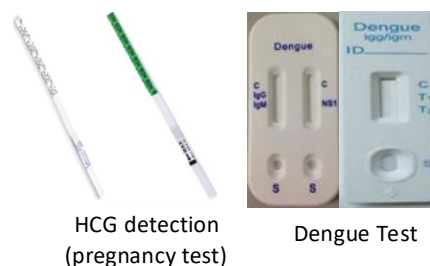


Figure 2: Brand to brand variation in colour and geometric shape. HCG: Human Chorionic Gonadotropin.

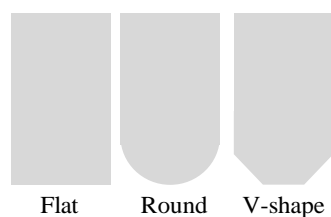


Figure 3: Variation in well shape

A colourimetric test provides decisive analysis of the present elements or concentration of the chemical compound facilitated by a colour agent. There is a wide range of pathological and colourimetric tests. The assay types, whether for a pathological test or not, can be defined in a number of ways, such as based on time-point, a number of analytes, signal amplification method, type of the substrate and format of the result (Tania *et al.*, 2018). Even within the same test type, more variation could occur due to size, shape, material and colour component utilised to develop the test by different commercial brands. Such physical variations are demonstrated in **Error! Reference source not found.-Error! Reference source not found.**. For example, in **Error! Reference source not found.**, the first picture shows an assay capable to test pH and nitrite in the water sample (Lopez-Ruiz *et al.*, 2014). The next picture is of a litmus test, which is one of the oldest forms of pH indicator, used to test materials for acidity. The assay shown in the last picture within **Error! Reference source not found.** is a mixture of several indicators that can evaluate how strong a given acid or base is, in a semi-quantitative manner.

The current practice to detect the assay type is by visual knowledge or bare-eye observation. Some key information about the test could be provided by the manufacturer. However, such type of detection can be challenging for naive users, especially due to lack of adequate knowledge within the domain including existing diversity (such as **Error! Reference source not found.-Error! Reference source not found.**). The focus of this paper involves the transformation of such visual knowledge regarding colourimetric pathological test type and chemical detection using DL as an alternative to bare-eye observation as shown in Figure 4.

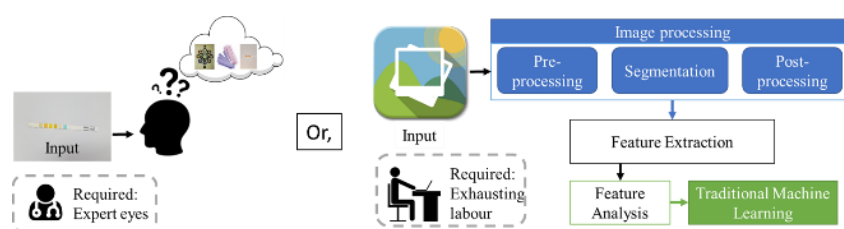


Figure 4: Two alternatives to DL-based approach for assay type detection

The automation of the bare-eye observation process by transforming the visual knowledge can aid in fulfilling the ASSURED criteria of ‘user-friendliness’ by offering more convenience and ease-of-use for less medically trained users, including at home and community settings. Such automation also enables portability of visual knowledge, which can fulfil ‘accessibility’ issue within ASSURED criteria.

The physical variabilities within the tests (such as **Error! Reference source not found.-Error! Reference source not found.**) are also worth exploring while developing an automatic image-based system to better manage the parameters involved. Such an understanding (Tania *et al.*, 2019) will be useful in order to develop universal intelligent colourimetric tests. The conventional quantitative colour related data extraction using simple analytic model suffers from various challenges such as loss of information during data compression, methods to counterbalance the loss of information, requirement of extracting multiple analytical models to track each colour change, inter-inoperability, lack of robustness, etc. some of which are indicated in (Guyon, Elisseeff and De, 2003; Solmaz *et al.*, 2018).

Therefore, machine learning is more preferred for such cases, especially when the number of independent input parameters increases (Koydemir *et al.*, 2015; Karlsen and Dong, 2016; Kim *et al.*, 2017).

Table 3: Variability regarding imaging (Tania *et al.*, 2019)

Premise	Parameter	Specifics
Camera parameter	Resolution (MP)	Low: 3.2 (Cooper <i>et al.</i> , 2012) High: 20.7 (Kim <i>et al.</i> , 2017)
	Sensitivity to light (ISO)	Varying from 50 (Alankus <i>et al.</i> , 2018) to 800 (Lopez-Ruiz <i>et al.</i> , 2014; Karlsen, 2018) Auto (Karlsen, 2018; Shabut <i>et al.</i> , 2018)
	Other parameters	Constant (Mutlu <i>et al.</i> , 2017) Auto (Shabut <i>et al.</i> , 2018; Solmaz <i>et al.</i> , 2018)
Camera to sample position	Distance	Low: 5cm (Yetisen <i>et al.</i> , 2014) High: 2 feet (Feng <i>et al.</i> , 2014)
	Exposure (Angle)	Parallel (Alankus <i>et al.</i> , 2018; Shabut <i>et al.</i> , 2018) Tilted (Karlsen, 2018)

On the other hand, the challenges within the image-based system include diverse real-world ambient conditions, variations in imaging conditions (**Error! Reference source not found.**), invariances (**Error! Reference source not found.**) etc. For an image-based system, as opposed to bare-eye observation, the data pre-processing steps would include image processing while utilising Traditional Machine Learning (TML) techniques. An illustration of such detection is shown in Figure 4. In such a case, one would require to separate the Region of Interest (ROI), which would be highly affected by the variabilities (such as in **Error! Reference source not found.**), some of the challenges apply to colourimetric detection as well.

Several approaches have been taken in the literature to circumvent these challenges and to produce a reliable colourimetric decision. The routes taken as an alternative to image segmentation are listed in **Error! Reference source not found.**

Table 4: Alternatives to image segmentation

Application	Additional hardware	Full image	Referencing	Cropping	QR code	Virtual guideline	Reference
H ₂ O ₂	No	No	No	Yes	No	No	Solmaz <i>et al.</i> (2018)
Urinalysis	Yes	Yes	No	No	No	No	Akraa <i>et al.</i> (2018)
	Yes	No	No	Yes	No	No	Rahmat <i>et al.</i> (2018)
pH test	No	No	No	Yes	No	No	Mutlu <i>et al.</i> (2017)
Saliva alcohol	Yes	No	No	Yes	No	No	Kim <i>et al.</i> (2017a)
Blood test	Yes	No	No	Yes	No	No	Kim <i>et al.</i> (2017c)

Water test	No	No	No	Yes	No	Yes	Sicard et al. (2015)
Wet-chemicals	Yes	No	No	No	No	Yes	Vashist et al. (2015)
Alkaline phosphate	Yes/No	No	Yes	No	No	No	Yu et al. (2015)
Blood test	Yes/No	No	No	No	Yes	Yes	Feng et al. (2014)

Each of these approaches (**Error! Reference source not found.**) has its benefit and downside. Inclusion of additional hardware would mean a contradiction to the criteria of ‘equipment-free’ within ASSURED criteria. Also, the addition of hardware can increase cost and reduce ease-of-use. As evident in **Error! Reference source not found.**, the existing literature often compromises automation by expecting the user to manually crop the image for better recognition of ROI.

An alternative approach to ease the process of test type detection could be the utilisation of virtual overlays (Figure 5). Currently, virtual plates are frequently used in the literature and also by the commercial applications to detect the sample, or the ROI, easing the traditional image processing steps (Alidans srl, 2015; Enzo Life Sciences inc., 2015; Vashist *et al.*, 2015; Sicasys Software GmbH, 2017). Such utilisations restrict the user with specific dimensions of the assay and sample positioning. Thus, these works lack in capacity to deal with assay-to-assay variations (**Error! Reference source not found.**-**Error! Reference source not found.**). For example, instead of intelligent sensing, the Spotxcel® Reader (Sicasys Software GmbH, 2017) uses a virtual plate which can be laid over the assay plate image. The application expects the wells to be aligned with the virtual plate. The user is required to match the corner and centre wells with the grid. The virtual plate or grid can be scaled and rotated. However, aligning the wells with the grid requires some image capturing skills, which reduces the ease of use. The software developers also acknowledged the limitations of their image processing scheme (Sicasys Software GmbH, 2017).

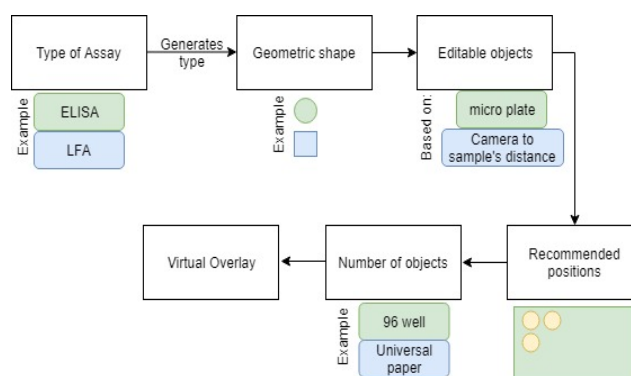


Figure 5: High-level architecture of the proposed virtual overlay based on the type of the assay

Therefore, one may adopt an intelligent virtual plate as proposed in (Tania et al., 2019). However, this paper aims to explore the strength and capabilities of DL algorithms in detecting the test type as well as to provide a colourimetric classification, rather than using any augmented reality as shown in **Error! Reference source not found.** Although machine learning has been investigated for colourimetric tests in the past (Feng *et al.*, 2014; Wang *et al.*, 2016; Kim *et al.*, 2017; Mutlu *et al.*, 2017), advanced machine learning techniques such as DL has not been explored for such cases yet which has the higher potential to produce an enhanced outcome.

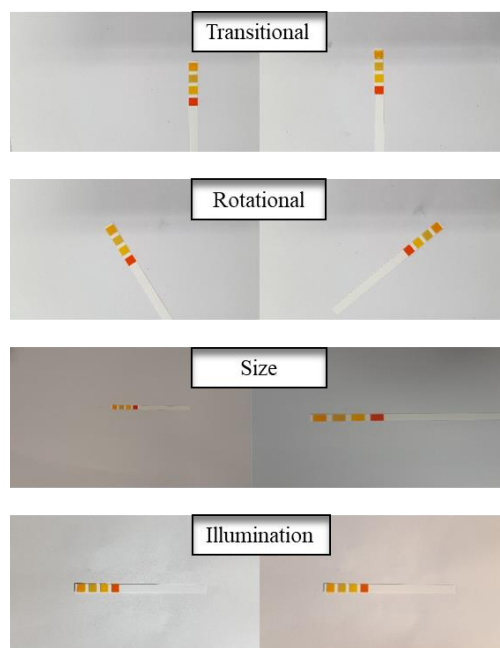


Figure 6: Invariances

The physical variations, demonstrated in **Error! Reference source not found.**, can play a significant role in an image-based system, which can be mitigated to a certain extent through deep neural networks (Tania *et al.*, 2018). One of the advantages of using DL algorithms in colourimetric pathological test type and chemical detection is a mechanism to deal with many of the variabilities. DL techniques for computer vision received a great response recently due to its capability of identifying objects even if it appears in a different form. For example, an apple is an apple whether red or green; a cat is a cat even if it tries to hide in a basket. This ability to recognise an object despite its different attire or appearance is called invariance. The invariance can be regarding position, rotation, size and even illumination, which is one of the most challenging issues for systems using traditional image processing algorithms. The diversities caused by invariances, along with other variabilities, can induce more hardship during handcrafted feature engineering while utilising TML. Few examples are shown in **Error! Reference source not found.**, which can be more convincingly handled by the pre-trained models that justify the use of DL algorithms in this work.

Figure 4 shows two alternatives of DL-based assay type detection; a DL-based approach can automate both the processes shown in the figure. Manual feature engineering, as shown in Figure 4, is a labour extensive process, limited by human time constraints and imagination. Moreover, the wide range diversity among assays (Tania *et al.*, 2019), variability due to ambient condition (Shabut *et al.*, 2018), variations due to manufacturers (**Error! Reference source not found.**), imaging condition (**Error! Reference source not found.**), variations caused by user behaviour and preference (**Error! Reference source not found.**) can be discouraging for the traditional image-based approaches, including TML, restricting the literature to rely on the human eyes, rather automating the process of the assay type detection. To the best of our knowledge, there is no existing work in the literature that detect the assay using DL techniques. Calibration techniques such as Yetisen *et al.* (2014) is often used to detect the assay or the sample, which can be time-consuming. The use of ‘control’ to detect the assay type has also been reported (Sicard *et al.*, 2015). The major advantage of DL techniques as demonstrated in this paper over the existing work using traditional approaches is the scale of flexibility to be deployed on similar assays.

3. Method

3.1 Dataset Preparation

This work used the same dataset as described in (Tania et al., 2019). The dataset for test type detection contains 1600 samples or images, with an equal distribution of two different assay type, as illustrated in Figure 7.

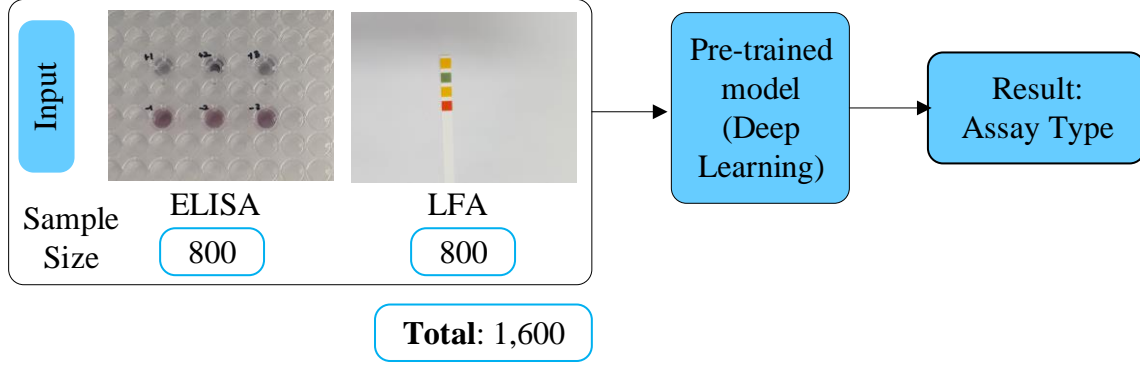


Figure 7: Framework for assay type detection using inductive transfer learning

The pathological test of the Gold Nanoparticles (AuNP) based plasmonic ELISA for TB-antigen specific antibody detection, referred to as the TB-test, is one of the case studies for assay type detection in this paper. The description of the sample preparation for the TB-test is provided in (Abuhassan *et al.*, 2017; Tania *et al.*, 2017; Bakhori *et al.*, 2018). However, the development of biosensor is not the focus of this work.

The other assay type utilised in this paper is pH indicator universal test strips, ranging from 3.0 to 10.0. It is a widely used dry-chemical based LFA, which is stable at room temperature and suitable for field operations as well. More details regarding the sample preparation procedure can be obtained from (Hoque Tania, 2018).

Both of these assay types are visually different in terms of size, shape and colour. These assays are also different due to the materials and functionality. The images of both assay types were captured by using iOS and Android-based mobile devices, and illumination conditions also varied greatly to create a robust dataset.

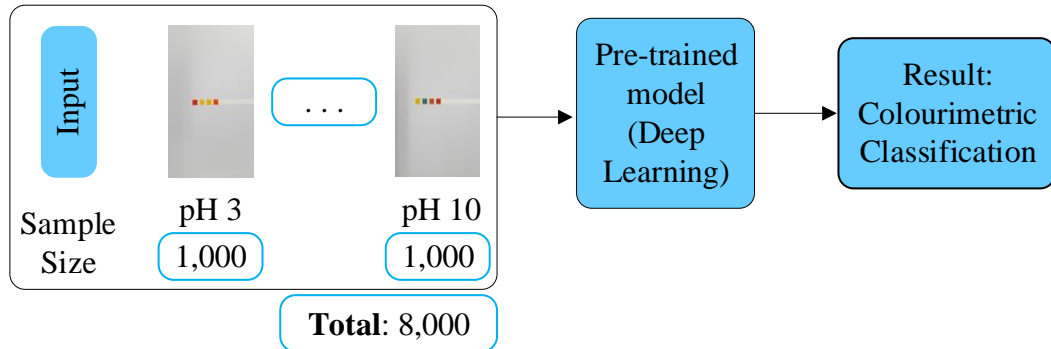


Figure 8: Framework for colourimetric classification using inductive transfer learning

The example of LFA was also chosen for the colourimetric classification using DL with a dataset of 1000 samples per class (Figure 8). The rationale for choosing this example is widespread usage, accessibility, suitability to be used within community settings, etc.

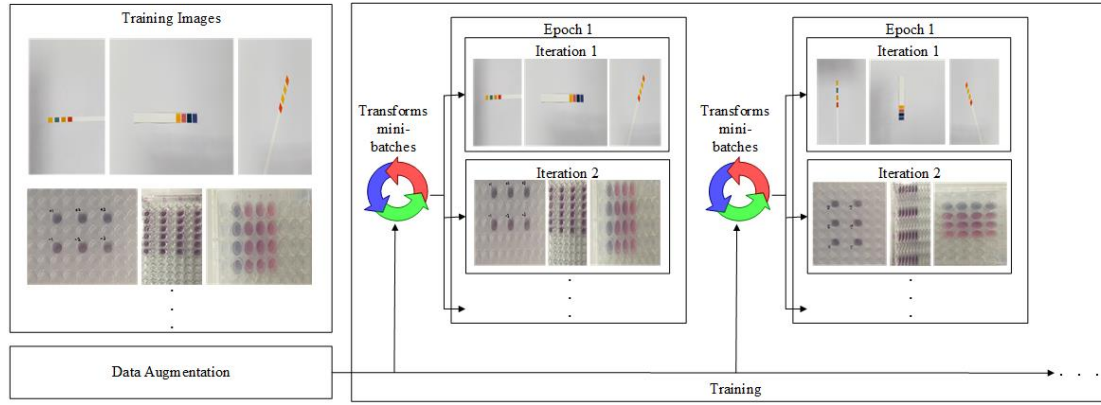


Figure 9: Example of data-augmentation for assay type detection.

Data augmentation techniques were utilised for both cases, i.e. dataset for assay type detection and colourimetric classification to manage the data partitioning for training, validation, testing, and prediction as illustrated in Figure 9, depicted from (MATLAB & Simulink, 2018b). In this way, the models were prevented to memorise the exact specifics of the training image that aids in reducing the overfitting problem.

3.2 Deep Learning-based Models

In this paper, the state-of-the-art DL models for visual recognition problems were critically explored to recognise the test or assay type utilising a Windows Intel(R) Xeon(R) Gold 6130 CPU at 2×2.10 GHz, 96GB with x64-based processor. The models were trained using MATLAB 2019b on a single Graphical Processing Unit (GPU).

The AlexNet (Krizhevsky, Sutskever and Hinton, 2012) is trained on more than a million images. It contains five convolutional layers and three Fully Connected (FC) layers which are marked in Figure 10. The concept of convolution is well described in Ng (2018).

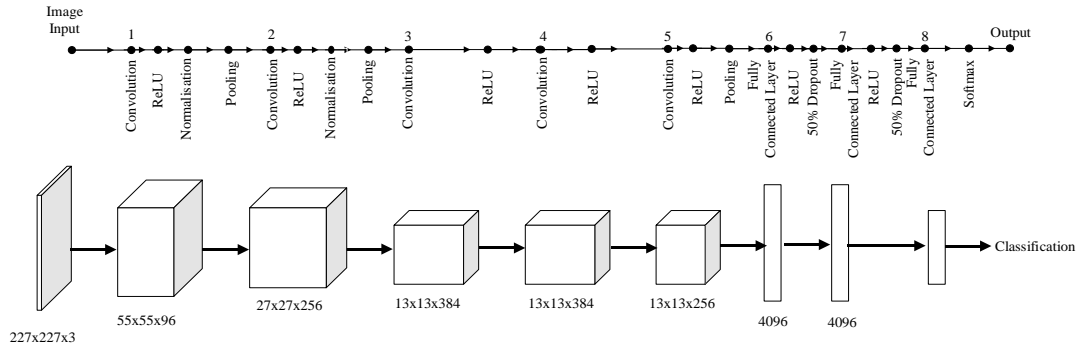


Figure 10: Eight layers of AlexNet

The fully connected layer utilises the output of the convolutional, activation or pooling layer (Scherer, Müller and Behnke, 2010) as the input volume and provides an N-dimensional vector, where N= number of classes. In the case of the softmax approach, FC considers the output of the previous layer and estimates which of the high-level features of the previous layer are the mostly correlated features for a specific class. By multiplying the input and the weight matrix and adding a bias vector, FC seeks to attain the correct probability for the classes. Using AlexNet in Figure 10, final layer, $fl=25$ and n^{th} layer before the final layer, $nl=2$ ($n \in \mathbb{Z}^+$; $n < fl$). AlexNet uses Rectified Linear Activation or ReLU function (Nair and Hinton, 2010).

In the Large-Scale Visual Recognition Challenge (ILSVRC2012), the AlexNet, containing two parallel convolutional neural network lines, showed 15.3% Top-5 error rate. The Top-5 error rate is the percentage of test examples for which the correct class was not in the top 5 predicted classes.

Few other pre-trained models were also explored in this work such as two versions of Inception (Szegedy *et al.*, 2015, 2016) and residual connections based ResNets (He *et al.*, 2016) allowing us to train the model with less number of samples and faster computation.

The inception layer aims to scan a bigger area while retaining a fine resolution for small information of the images. The network architecture (Szegedy *et al.*, 2015, 2016) presented more freedom regarding the specifics of the convolution. Each inception module can perform a number of parallel convolutional computations and concatenate. With a hope that a series of Gabor filters of different sizes will be able to provide better management of multiple object scales, the parallel convolution of different sizes starts from the most accurate detailing, i.e. 1×1 , continues to the bigger ones (Figure 11).

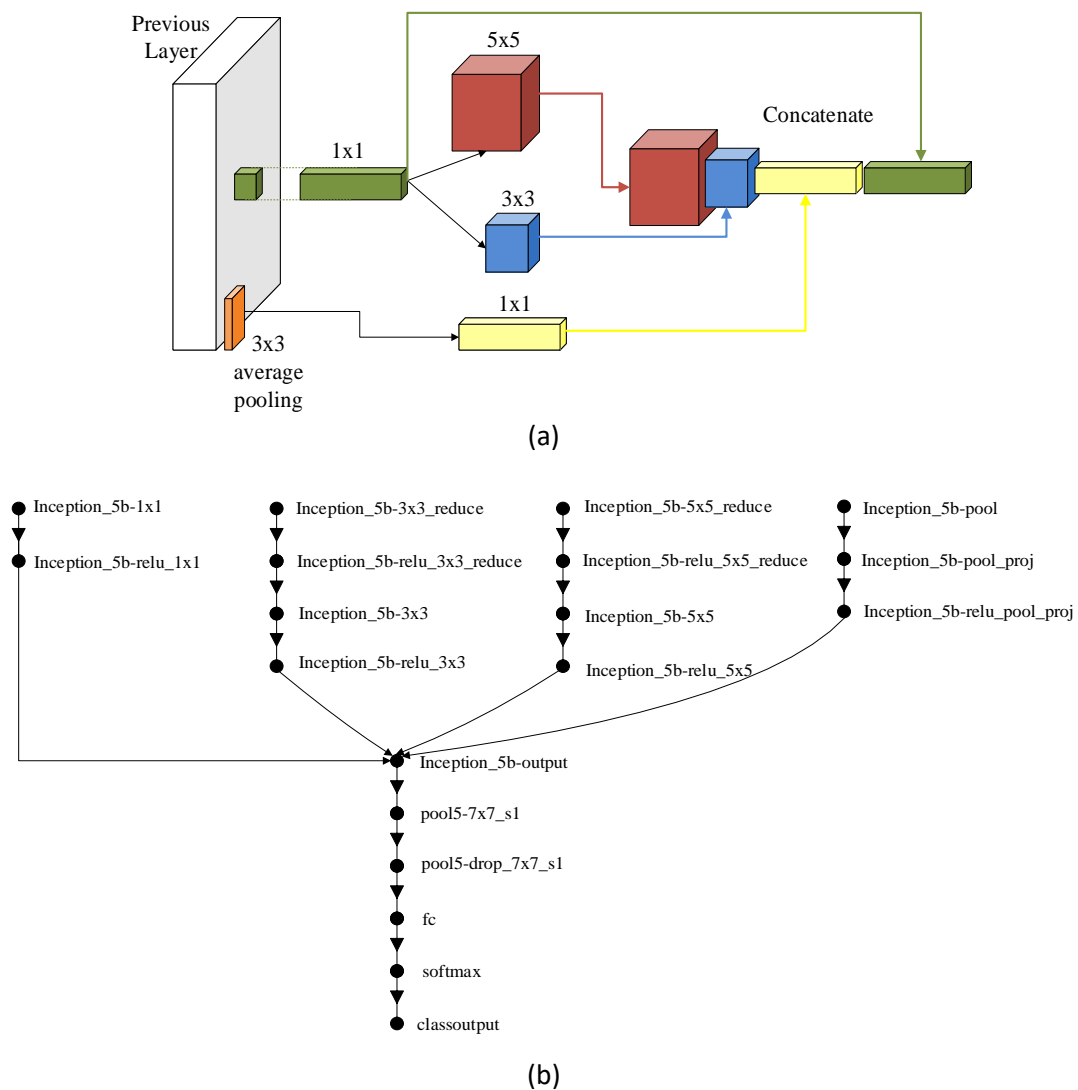


Figure 11: Architecture of GoogLeNet. (a) Basic idea of inception modules, (b) Final layers of fine-tuned GoogLeNet

A 1×1 convolution maps an input pixel along with the associated channels to an output pixel, not looking at anything around itself. It is often used to reduce the number of depth channels since it is often very slow to multiply volumes with extremely large depths. Theoretically, the neural network can 'choose'

which input 'colours' to look at while using this mapping, instead of multiplying everything, which would be extensively expensive.

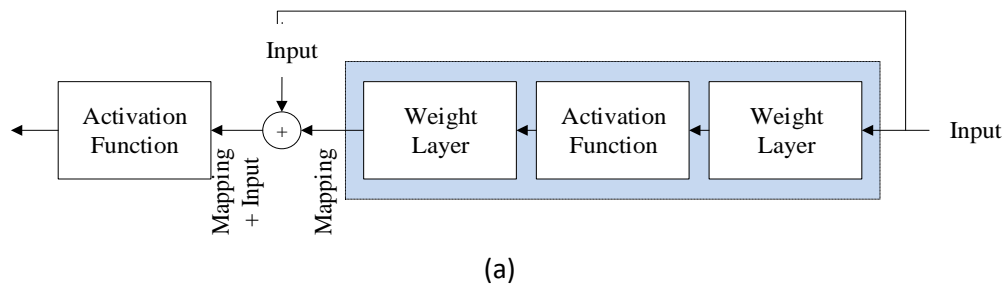
Unlike the traditional Gabor filters, the filters of the inception layers are learnable. In order to prevent overfitting due to a huge number of parameters involved, the Inception exploited bottleneck approaches while rebuilding the inception module with more non-linearities and fewer parameters, followed by a max-pooling layer to summarise the content of the previous layer. The results obtained are concatenated one after the other and passed on to the next layer (Figure 11).

To fine-tune the model, in the case of inception modules based GoogLeNet, the final three layers of the network have to be replaced to retrain the model with our dataset. These three layers are 1000 fully connected layer called 'loss3-classifier', softmax and the output layer. These layers embrace the mechanism to train the model from the extracted features into class probabilities and labels. The final layers can be visualised in Figure 11.

In the case of an advanced version of GoogLeNet, i.e. Inception-v3, the final three layers are replaced and connected to the 313th connection that performs average pooling.

This paper also explored ResNets. The ResNet (He *et al.*, 2016) investigated that if larger function classes contain the smaller ones, then can it be guaranteed that increasing them strictly increases the expressive power of the network. The basic idea of ResNet includes the identity function as one of the elements while progressing through the network (Figure 12). Therefore, when a newly added layer is trained into an identity mapping, $f(x) = \text{input} (= x)$, the new model is supposed to be as effective as the original model.

The new model can potentially provide a better solution to fit the training dataset, hence the added layer may aid in reducing training errors. Using ResNets, the gradients can flow directly through the skip-connections (Figure 12) backward from later layers to initial filters.



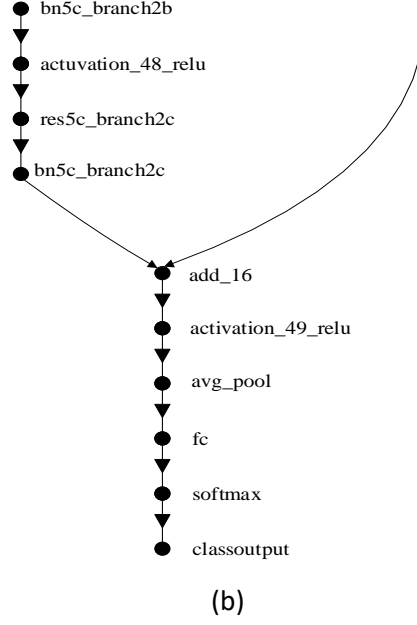


Figure 12: Architecture of ResNet. (a) Basic idea of residual network, (b) Final layers of fine-tuned ResNet-50

Similarly to GoogLeNet, the final layers have to be replaced to fine-tune the model of 50 layers deep ResNet-50 (Figure 12) and 101 layers deep ResNet-101.

3.3 Fine-tuning the Models

Despite the advantages of DL techniques in advanced analysis and feature extraction in supporting the non-vision experts (Krizhevsky, Sutskever and Hinton, 2012; Szegedy *et al.*, 2015; He *et al.*, 2016), it is a data-hungry, resource-demanding technology, which contradicts with our attention towards the ASSURED criteria (Hoque Tania *et al.*, 2020).

Table 5: Specifications of the pre-trained models

Model	Model size (MB)	Image Input Size	Depth	Frozen layers (up to)	Parameters (Millions)	Network
AlexNet	227	227×227	8	0	61.0	Series
GoogLeNet	27	224×224	22	10	7.0	DAG
Inception v-3	89	299×299	48	18	23.9	DAG
ResNet-18	44	224×224	18	5	11.7	DAG
ResNet-50	96	224×224	50	10	25.6	DAG
ResNet-101	167	224×224	101	4	44.6	DAG

DAG: Directed Acyclic Graph

Alternatively, the success stories of ImageNet project (ImageNet, 2016), trained on million images, has opened the door for faster implementation of DL via inductive transfer using limited resources (such as smaller dataset), especially when there is a substantial amount of large data and similarity between the tasks is also high. Therefore, this paper utilised the following models as described in Table 5 using Stochastic Gradient Descent with Momentum (SDGM) as for their solver training network (Murphy, 2012).

$$\theta_{l+1} = \theta_l - \alpha \nabla E(\theta_l) + \gamma(\theta_l - \theta_{l-1}) \quad (1)$$

where θ = parameter vector, l = iteration number, $E(\theta)$ = cost function, initial learning rate = α and momentum $\gamma=0.9$ (Brownlee, 2018). Equation 1 aids the solver networks of Table 5 to accelerate gradients vectors in the right directions, ensuring a faster convergence. Gradient clipping was utilised to prevent gradient explosion.

$$E_R(\theta_l) = E(\theta_l) + \lambda\Omega(w) \quad (2)$$

A weight decay was added to the cost function (Equation 2) to reduce overfitting (Bishop, 2006; Murphy, 2012), where w = weight vector, λ = regularisation co-efficient and $\Omega(w)$ = regularisation function. $E_R(\theta_l) = 1 \times 10^{-4}$ (Hoque Tania, 2018).

One of the benefits of pre-trained models is, the pre-trained weights can be better than randomly initialised weights. Therefore, one should aim at retaining the essence of the traits of the original models, so that the pre-trained weights are not highly altered instantly. Therefore, the common practice is to use an initial learning rate ten times smaller than the actual model. Although a smaller learning rate would elongate the convergence time, this hyperparameter can assist in tracking all the local minima by controlling the weights-adjustment with regarding the loss gradient. In this paper, $\alpha = 0.001$ for AlexNet, whereas $\alpha = 0.0001$ for the rest of the models.

The features extracted in the initial layers of DL-based models (such as Table 5) focus on the basic information that can be extracted from the images. For example, the same edges and arcs that form the image of a dog or a cat in the original dataset (ImageNet, 2016) can make up a well in Figure 7: Framework for assay type detection using inductive transfer learning. Therefore, it is possible to leverage the knowledge of meaningful general features at initial layers to be used for another problem, as the generic information often would be the same, whereas the later layers are more specific to the particular task. Based on the problem of this paper, some layers were chosen to be frozen, which are indicated in Table 5.

The size of the dataset and the similarity between the datasets with the original dataset used to train the model can help to repurpose the pre-trained models, which is the rationale for choosing the rectangular box in Figure 13 as the training method. Figure 13 is inspired by Marcelino (2018).

The datasets utilised in this paper are relatively small, containing 800-1000 images per class as shown in Figure 7: Framework for assay type detection using inductive transfer learning and Figure 8. For a small dataset with higher similarity with the original dataset, one could freeze the convolutional base, leaving only the last fully connected layer – which is not the case in this paper. The datasets of this paper are different than the original, i.e. pre-trained dataset (ImageNet, 2016). Therefore, it is important to find a balance between the layers to train and to freeze to avoid under- or overfitting.

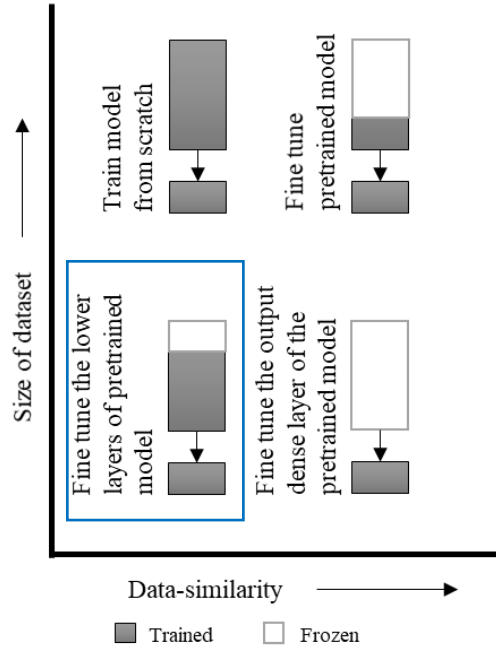


Figure 13: Thumb-rule to fine-tune pre-trained models.

The learning rate in the frozen layer is zero. Freezing initial layers can prevent those earlier layers from overfitting to the presented dataset, whereas freezing all the layers will prevent the models to learn anything meaningful, as the network would not update the parameters of the frozen layers. Therefore, this paper freezes the initial k layers of the pre-trained models (Table 5) and train the later $n - k$ layers using the concept of the rectangular box in Figure 13. The first k layers form the initial 'stem' of the networks – which was our selection criteria to choose the number of frozen layers in Table 5. One of the direct impacts of such a training method is on the computation time as, in contrast to backpropagating and updating the weights of all the layers of the network, we now needed to backpropagate the gradient and update the weights of the last $n - k$ layers. More detailed concept of the rationale behind the thumb-rule can be well perceived from the literature such as Marcelino (2018) and Mathisen *et al.* (2019).

Using these pre-trained models, i.e. AlexNet, a number of Inception and ResNet models, with the aforementioned parameters, this paper at first aimed at providing the classification of the assay using the case study of Enzyme-Linked ImmunoSorbent Assay or ELISA (Shabut *et al.*, 2018) and Lateral Flow Assay or LFA (Hoque Tania *et al.*, 2020). The paper was further extended to provide an interclass classification of the assay using the example of colourimetric classification of the LFA (Hoque Tania *et al.*, 2020).

The training cycle of AlexNet had 340 iterations, distributed among 20 epochs during the assay type classification, and 1740 iterations, equally divided among the similar number of epochs during the intraclass classification. An epoch is one learning cycle where the learner sees the whole training dataset. Since one epoch is too big for computation at once it is divided into smaller batches. One epoch being the entire dataset passed forward and backwards through the network only once, and gradient descent being an iterative process that optimises the learning rate— updating the weights with one epoch during training is not enough. Hence, the training cycles require an adequate number of epochs. Due to the nature of the networks, Inception and ResNet models had 672 iterations, distributed among 6 epochs during the assay type classification, and 3360 iterations, equally divided among the similar number of

epochs during the intraclass classification. As the dataset of colourimetric classification is larger than the assay types', the number of iteration and iteration per epochs were understandably higher for the former.

4. Result and Discussion

4.1 Assay Type Detection

All the models listed in Table 5 showed 100% accuracy to classify the type of the assay (Figure 14). Therefore, convergence time, memory occupancy and network complexity were taken into account to select which model to utilise (Table 5, Figure 14 and Figure 15).

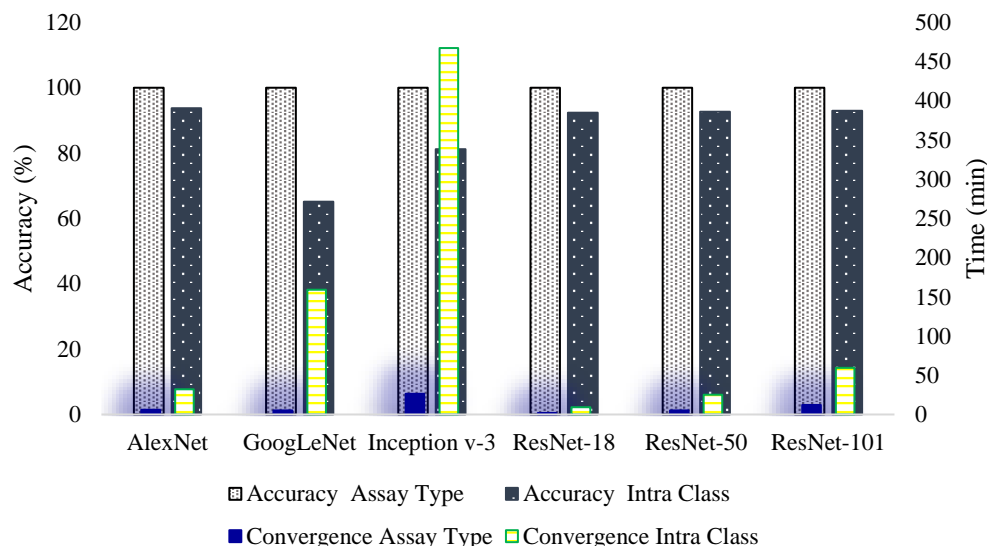


Figure 14: Comparative performance in terms of accuracy and elapsed convergence time for assay type and colourimetric classification

In addition to accuracy, Figure 14 also portrays the comparative position of the models (AlexNet, Inceptions, ResNets) in terms of elapsed convergence time for assay type and colourimetric classification using the entire dataset, which can also be perceived from the computation method, as discussed in Section 3.

As shown in Section 3.2 and also in Table 5, the layers of AlexNet are arranged one after the other, forming a series network, whereas the layers of the rest of the models hold more complex architecture using Directed Acyclic Graph (DAG). However, in terms of memory occupancy, AlexNet is not the smallest one is the size (Table 5 and Figure 15). The network, as well as the performance, can be optimised for some DAG network-based models such as ResNet-18. ResNet-18 provided 100% accurate classification of the assay types taking around 2 minutes to train on 1600 images occupying less memory space. Therefore, ResNet-18 is a preferred choice of pre-trained model for the task of classification of assay types.

4.2 Colourimetric Detection

The intention behind using pre-trained models was to determine the type of assay, where the type of assay signified the class label. This paper extended the experiments on the pre-trained models for intra-class classification of the assays, i.e. colourimetric classification. The LFA example was chosen for this extension to evaluate the scope of using pre-trained models for a complete colourimetric classification.

Using 1000 images for each pH in between 3.0 to 10.0, the best accuracy was achieved by AlexNet (Figure 14). In Section 4.1, the classification accuracy was found to be 100% for the assay type classification, whereas, for colourimetric detection, it is over 92% for AlexNet, ResNet-18/50/101.

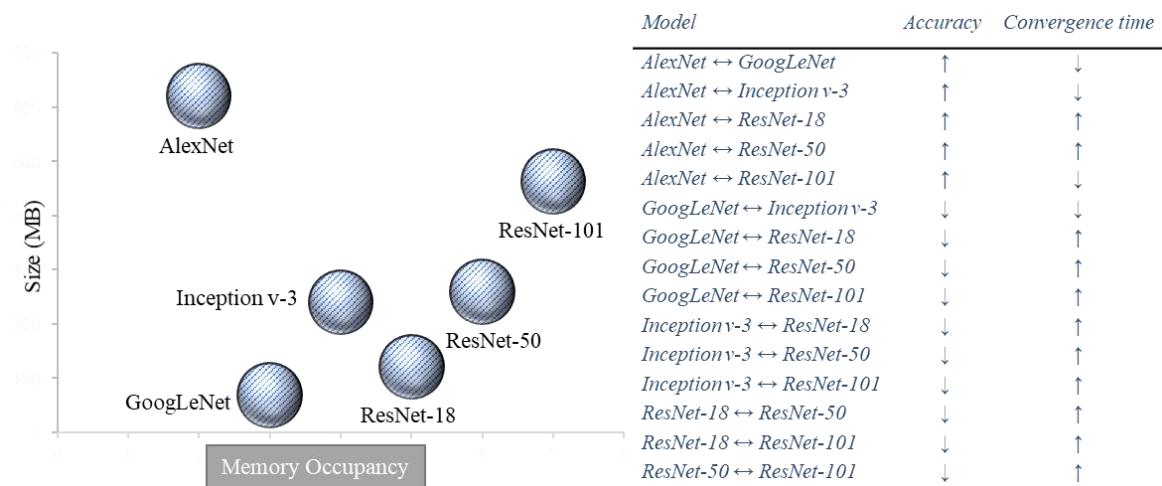


Figure 15: Comprehensive comparative performance for colourimetric classification

Figure 15 shows a comprehensive comparison among the models, especially for the intra-class classification. The statistical results obtained by a two-tailed t-test at a 0.05 level of significance are also given in the figure. The result of Algorithm 1 ↔ Algorithm 2 stated as “↑”, “↓”, or “~” can be read as- Algorithm 1 is significantly better than, worse than, or statistically equivalent to Algorithm-2, respectively. For example, AlexNet performed significantly better than GoogLeNet, in terms of accuracy, however, it took significantly longer time to be trained than the later. As the accuracy of intra-class classification was not as high as the assay type classification, this issue was further investigated.

As AlexNet showed the best performance for the colourimetric classification of pH levels using universal pH indicator paper, i.e. 93.75%, the output of the convolution layers were critically investigated. Figure 16 shows an arbitrary input image of pH 7 as an example.

The 1st convolutional layers of AlexNet are consist of 96 11x11x3 convolutions with [4 4] stride and [0 0 0] padding (Krizhevsky, Sutskever and Hinton, 2012). The output of activations can be represented by a 3D array, where the 3rd dimension is a representation of the channel of the 1st convolutional layer. As activations are colour blind, subsequent to normalisation and scaling, the output of the activation is shown as grey images in Figure 16, where minimum activation=0 and maximum activation=1. For better visualisation, Figure 16 only shows the strongest activations at each convolution layer. The features at the final fully connected layer, i.e. 23rd layer of AlexNet are also shown in Figure 16 using Deep Dream Images (Mordvintsev, Alexander; Olah and Tyka, 2015; Mordvintsev, Alexander; Olah, Christopher; Tyka, 2015; MATLAB & Simulink, 2018a). With an increase in the number of pyramid levels as well as the iterations per pyramid level, enlarged output with strong activation at the 23rd and 24th layer are also shown in Figure 16. These detailed output images of each layer failed to provide any significant information regarding the low classification accuracy. These layers of AlexNet were supposed to search for only colours of each block of the pH test strip and block to block colour difference. Considering the strong activations at the convolutional layers, it can be estimated that the most massive activations in the deeper layers may not be the point of interest. The point of interest may have been buried in some other channels. Furthermore, the deeper layers of AlexNet in Figure 16 implicated both positive and negative activations that was later rectified in the ReLU layers.

Among the pre-trained models, Inception based network showed poor performance- GoogleNet achieved only ~65% accuracy. Despite the use of deeper layers, Inception v-3 reached up to 81.25% accuracy. Further exploration into the convergence pattern could not reveal any indication if only a longer training cycle or a bigger dataset would have offered a better accuracy or not.

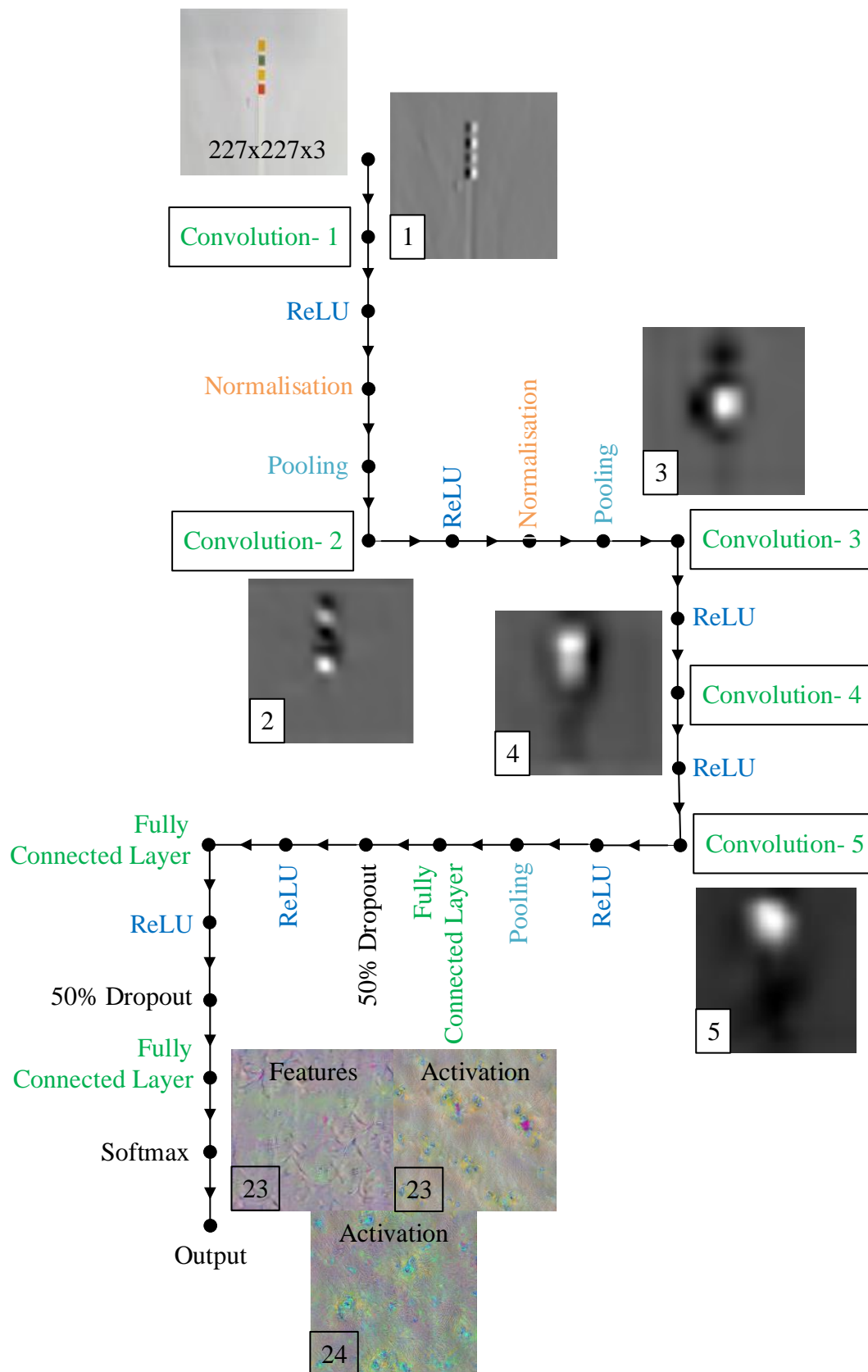


Figure 16: Layer by layer investigation of the performance of AlexNet for intra-class classification

ResNets attained similar accuracy as AlexNet, as shown in Figure 14. Originally, ResNet gained popularity due to the counter-intuitive method of randomly dropping layers during training, while using

the full network in testing, circumventing the struggle of longer training period for using deeper layers in other models such as Inception (Huang *et al.*, 2016). Analysing the performance in Figure 14, it can be seen that deeper layers did not significantly improve the performance of ResNets either.

As compared to the assay type detection, the networks required considerably higher time (Figure 15) to be trained on due to the computational complexity of the algorithms for the colourimetric classification as shown in Figure 14. Understandably, the increase in the dataset, as well as the number of classes, affected the training time. Similar to assay type detection, ResNet-18 was found to be the fastest model to be trained on.

Based on the accuracy (Figure 15), AlexNet was chosen to be further explored to investigate the accuracy attained and the scope of further improvements. The convergence rate (Figure 17) indicates a longer training cycle may improve the performance of the network, which would also expose the model to the risk of overfitting problem. The need for a significantly large dataset and longer training cycle would effectively require more enhanced resources, e.g. higher computational power, which contradicts with our research goal.

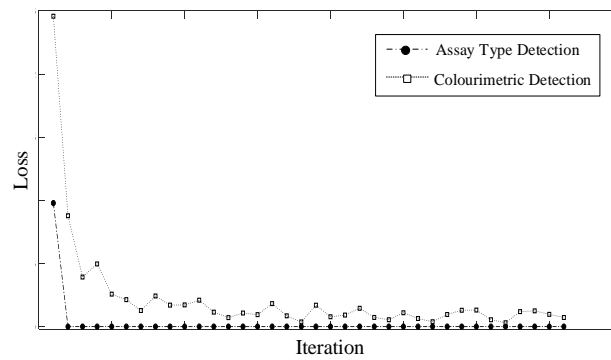


Figure 17: Convergence of AlexNet for assay type detection and colourimetric detection

However, to test our hypothesis that a larger dataset may improve the performance, in the absence of a larger dataset, we have downscaled the size of the dataset, while maintaining equal class-balance, and tracked the performance in terms of accuracy with respect to the increase in data. The dataset split ratio for training and testing were also varied for both datasets, for the same purpose, i.e. to observe the model's ability to learn from more data.

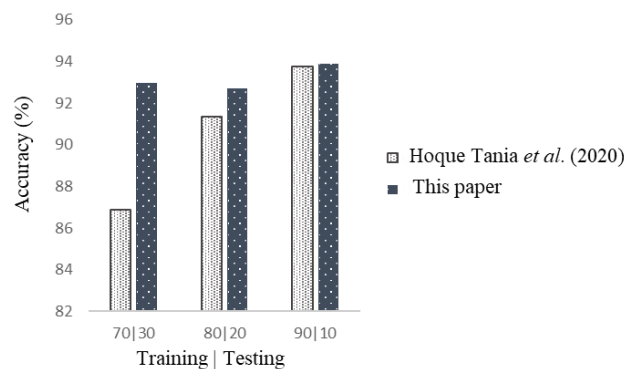


Figure 18: Increase in accuracy with respect to increment in data size

The dataset was downsized up to the dataset of Hoque Tania *et al.* (2020), consisting of 65 images for each of the 8 chosen pH levels. The classification accuracy for Hoque Tania *et al.* (2020) dataset, using 70% of them for training, was 86.86%. The result in Figure 18 shows, the AlexNet was able to learn more from new data, as the accuracy increased up to ~7% when more data were added.

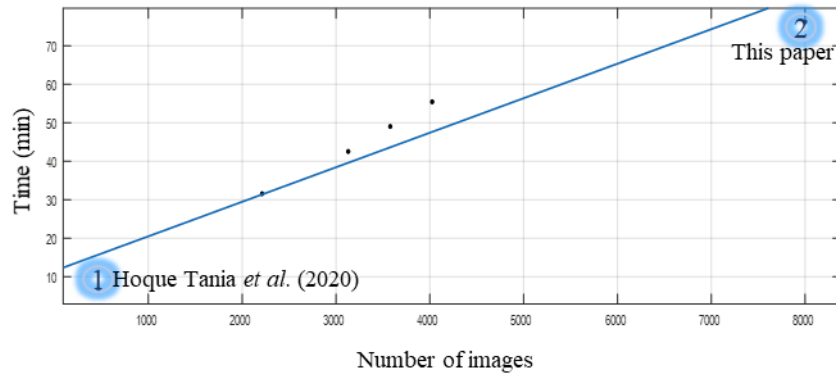


Figure 19: Increase in elapsed time with respect to increment in data size

More data, of course, require more computation, resulting in higher computation time, which can be observed in Figure 19. Interpolation techniques have been utilised for better visualisation of the trend (Figure 19).

On the other hand, the intra-class classification was conducted for LFA only, not ELISA due to the following reasons.

- For proof-of-concept, we found only one intra-classification would be sufficient because the intra-class classification accuracy obtained by the pre-trained models was unsatisfactory. Therefore, the approach would have had to be rejected for the final colourimetric classification, despite the outcome of the case study of ELISA.
- The case study of LFA is a stable assay. The dataset possesses consistency. A large number of images were fed to the network. The classification accuracy was still poor. The case study of ELISA contains less stable samples. Moreover, it is sensitive to temperature and geo-location. Therefore it would be more difficult to understand and explain the underlying reasons for the hypothetical poor performance of the pre-trained models.
- The analysis of the classification accuracy for LFA dataset indicated towards the requirement of more samples. The available ELISA dataset was already smaller than the LFA dataset. Therefore, the number of images would be insufficient for these data-hungry pre-trained models.

4.3 Smartphone-based Detection

For both assay type and colourimetric detection, in addition to the desktop, the prediction was delivered on the mobile platform. The models were trained offline as mentioned in Section 3, before deploying them on the mobile platform.

Although the proposed method in this paper involves pre-trained models with initial frozen layers training these CNN-based algorithms mostly rely on powerful computation resources and large amounts of training data, which can be more convincingly achieved using high-performance CPUs, GPUs, etc. Thus, the DL-based models are heavy for mobile devices. Mobile devices can be restricted in terms of storage and processing capability. Most smart mobile devices also suffer from the power consumption issue with limited battery capacity. On the other hand, once the model is built or trained on a high-powered computer or a cloud, it can be deployed on a mobile device with much less computation power.

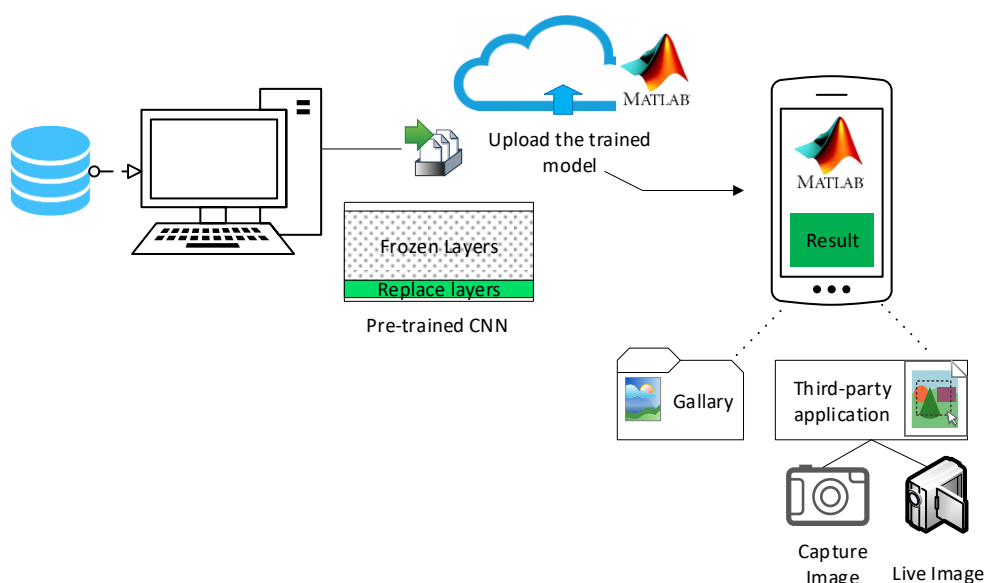


Figure 20: Mobile-enabled server-based pathological test type and chemical detection using deep learning

This paper utilised a third-party application that enables on-site image capturing on a server connected-MATLAB mobile. The server can take about a minute to establish a connection with a mobile phone, which may vary based on internet quality and mobile device. The result can be provided on a mobile platform as shown in Figure 20: Mobile-enabled server-based pathological test type and chemical detection using deep learning.

This paper utilised more affordable mobile devices, such as the Samsung Galaxy S6. Due to the limited storage and processing capabilities of mobile devices, the server was utilised to retain the accuracy of the system. Identical prediction accuracy was achieved on the server connected mobile platform as well. The prediction was delivered in real-time, however, taking slightly more time than the desktop environment, which supports the finding of Kim et al. (2017) that indicated that a server-based approach could make a colourimetric classification system two times slower.

4.4 Discussion

Based on the analysis of the results obtained by the pre-trained models, considering the accuracy, training time, memory occupancy, network structure ResNet-18 can be chosen to determine the type of test to assist a less medically trained person such as home users, aid workers, to begin a colourimetric-based pathological test without any prior knowledge about the nature of the assay (Figure 14).

The accuracies of the final colourimetric test by the pre-trained models were below 95%. The geometric shape of the assay and location of the coloured samples are easily distinguishable for these pre-trained models using the standard kernels with a reasonable amount of dataset. However, the intra-class classification would require the model to search for only the colours even when the rest of the geometric features are similar, and the occurrence of those colours are in the same location. The accuracy of the colourimetric classification may improve with a higher number of data, deeper layer or longer training cycle (Figure 18). Alternatively, better accuracy may result from building the model from scratch. Either of the routes would be computationally expensive (Figure 19).

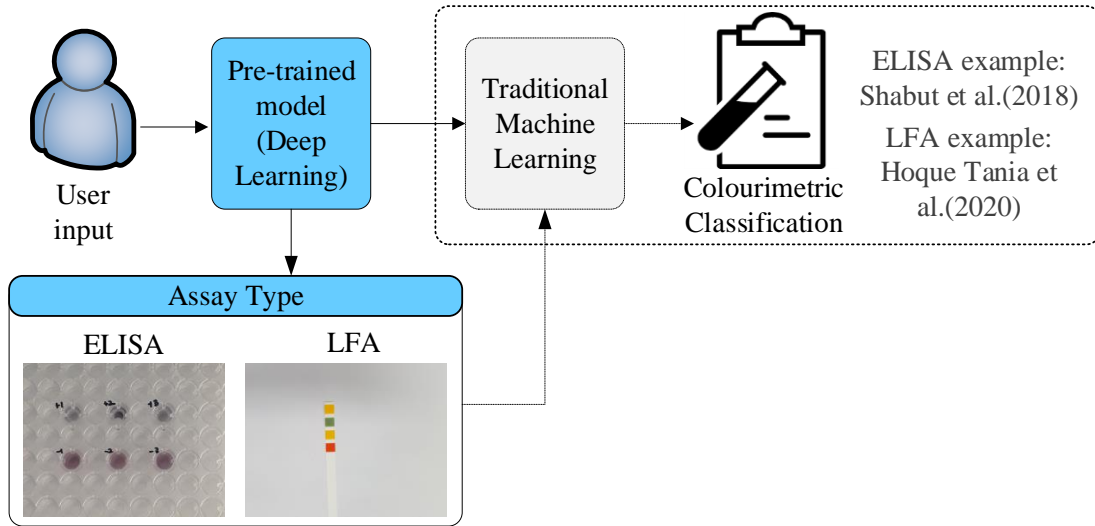


Figure 21: Proposed framework for assay type detection and corresponding colourimetric classification

Therefore, based on the aforementioned discussion, subsequent to the assay type detection, it is more logical to use a simpler machine learning model (Figure 21: Proposed framework for assay type detection and corresponding colourimetric classification) for the colourimetric classification, instead of building more deep layers which would require more processing capacity, memory size, larger dataset and more dependency on the cloud-based approach. In the literature, Hoque Tania *et al.* (2020) and Mutlu *et al.* (2017) utilised the same case study of LFA to perform colourimetric classification as this paper. Both of these reported studies vary in compatibility with the ASSURED criteria. Taking a different approach to feature extraction, both Hoque Tania *et al.* (2020) and Mutlu *et al.* (2017) settled for utilising Least-squares Support-Vector Machine (LS-SVM) to classify the pH levels, achieving 100% classification accuracy. The important considerations to perceive the comparative performance between the DL-based models in this paper and TMLs in the literature is shown in Table 6. Although the dataset utilised by Hoque Tania *et al.* (2020) and Mutlu *et al.* (2017) were much smaller in size, as compared to the dataset utilised in this paper, the colourimetric classification accuracy attained by the traditional approaches (Table 6) shows a higher prospect if a combination of DL and TML is utilised as proposed in our framework (Figure 21: Proposed framework for assay type detection and corresponding colourimetric classification).

Table 6: Comparative performance of colourimetric classification of the LFA

	Algorithm	Sample per Class	Number of Class	Invariances	Data Aug.	Prior Feature Extraction	Accuracy (%)	Reference
DL	AlexNet	1000	8	✓	✓	x	93.75	This paper
	GoogLeNet			✓	✓	x	65.17	
	Inception v-3			✓	✓	x	81.25	
	ResNet-18			✓	✓	x	92.38	
	ResNet-50			✓	✓	x	92.71	
	ResNet-101			✓	✓	x	92.98	
TML	LS-SVM	65		✓	x	✓	100	Hoque Tania <i>et al.</i> (2020)
	KNN			✓	x	✓	98.5	
	LDA			✓	x	✓	98.3	

LS-SVM	18	15	✓	x	✓	100	Mutlu et al. (2017)
--------	----	----	---	---	---	-----	---------------------

Based on the aforementioned discussion, this paper suggests (Figure 21: Proposed framework for assay type detection and corresponding colourimetric classification) to utilise pre-trained model-based DL algorithms for the assay type and to process the outcome using traditional machine techniques to produce the colourimetric decision as proposed in (Shabut *et al.*, 2018; Hoque Tania *et al.*, 2020). In this way, the image-based intelligent colourimetric test would provide more automation while processing the pathological test or colourimetric chemical detection while maintaining high accuracy and reliability, faster computation, and retaining more autonomy to users.

5. Conclusion

The colourimetric assays have been utilised for a wide range of applications ranging from environmental monitoring to disease diagnosis. This paper presented the transfer learning methods to build the CATC model which can be used by novice users such as patient or their caregiver to begin a colourimetric test at homecare settings, community settings, telepathology etc. To the best of our knowledge, no similar work has been reported until now.

Some of the mostly used pre-trained models such as AlexNet, GoogLeNet, Inception and ResNet were utilized in this paper for classification of the assay types. The use of the pre-trained model can help with faster training for similar applications with a relatively small dataset (<1000 samples per class). The use of DL techniques also aids in tackling physical variability such as brand-to-brand variations, illumination conditions, etc. Such an image-based intelligent system also enables convenient portability of the results that can be used in telemedicine, digital healthcare and m-health services.

Exploring the relevant pre-trained models, ResNet-18 was found to be the most suitable DL technique for detecting pathological tests in terms of low computational complexity, high accuracy, and low memory occupancy. Based on the current feasibility at resource-limited settings, our research suggests to utilise the pre-trained models for the test type detection, and then pass the outcome to a linear model using traditional machine learning algorithms to produce the colourimetric detection, enabling higher visual knowledge transformation and to make the overall process more intelligent, rapid and robust. Furthermore, the work has been extended by developing a low-cost handheld colourimetric assay detection device to determine assay types for various applications (Figure 20: Mobile-enabled server-based pathological test type and chemical detection using deep learning).

This work attained 100% accuracy for assay type detection using transfer learning approach, which can be deployed for any similar assays and applications. One of the key considerations of this paper was ASSURED criteria, by retaining ease-of-use and accessibility of the computational technology to the end-users. Therefore, this paper has not explored develop-model approach, which can be more data-hungry and resource-demanding.

However, our future work will utilise developed-model approach, while complying with ASSURED criteria for further advancement of the system as the pre-trained models achieved <95% accuracy in this paper for the intra-class colourimetric classification. In order to employ the developed-model approach, our future efforts will require to generate a larger dataset. Henceforth, we will aim at a further extension of samples per class and variability to increase the reliability of the system. Additionally, inducing more diversity in the assay will demonstrate and evident our claim which can ultimately result in a universal model for image-based colourimetric testing. In such a situation, consideration of the universal applicability can verify the adaptability of the presenting system as well. Integration of reinforcement learning can update the system based on environmental factors (e.g. geo-location, climate) and change in the pattern of the assay. Realising the potential of the presented system, future

work will also seek opportunities for commercialisation of the comprehensive framework as proposed in Figure 21: Proposed framework for assay type detection and corresponding colourimetric classification.

Acknowledgement

The dataset of this paper was collected as part of the doctoral research of the first author, funded by Erasmus Mundus Partnerships Action 2 “FUSION” (Featured eUrope and South asla mObility Network). Grant reference number: 2013-3254 1/001001. The authors extend their gratitude to Dr Khin Lwin and Dr Antesar Shabut for their support in the PhD project. The authors thank Prof Nor Azah Yusof and her team, Universiti Putra Malaysia for their support to collect the ELISA dataset. The original ELISA dataset was generated as part of the project named ‘TB-Test - A smart mobile enabled scheme for tuberculosis testing’, supported by British Council Newton Institutional Links and Newton-Ungku Omar Fund (Grant ID: 216385726). The authors also thank Dr Mohammad Najlah and Mr Paul Cotton, Anglia Ruskin University for their support to conduct the laboratory experiments on LFA.

References

- Abuhassan, K. J. *et al.* (2017) ‘Automatic Diagnosis of Tuberculosis Disease Based on Plasmonic ELISA and Color-based Image Classification’, in *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Jeju Island, South Korea: IEEE, pp. 4512–4515. doi: 10.1109/EMBC.2017.8037859.
- Akraa, S. *et al.* (2018) ‘A smartphone-based point-of-care quantitative urinalysis device for chronic kidney disease patients’, *Journal of Network and Computer Applications*. doi: 10.1016/j.jnca.2018.04.012.
- Alankus, G. *et al.* (2018) ‘Single-Image-Referenced Colorimetric Water Quality Detection Using a Smartphone’, *ACS Omega*, 3(5), pp. 5531–5536. doi: 10.1021/acsomega.8b00625.
- Alidans srl (2015) *AssayColor, Android App on Google Play*. Available at: <https://play.google.com/store/apps/details?id=com.alidans.assaycolor> (Accessed: 10 January 2017).
- Arnett, N. *et al.* (2016) ‘Inexpensive urinalysis test strips to screen for diabetes in developing countries’, in *2016 IEEE Global Humanitarian Technology Conference (GHTC)*. Seattle, WA: IEEE, pp. 589–596. doi: 10.1109/GHTC.2016.7857339.
- Arvaniti, E. *et al.* (2018) ‘Automated Gleason grading of prostate cancer tissue microarrays via deep learning’, *Scientific Reports*. Nature Publishing Group, 8(1). doi: 10.1038/s41598-018-30535-1.
- Bakhori, N. M. *et al.* (2018) ‘Immuno Nanosensor for Ultrasensitive and Affordable Naked Eye Detection of Tuberculosis’, *Sensors*, 18(6), pp. 1–10. doi: 10.3390/s18061932.
- Bhattacharjee, T., Jiang, H. R. and Behdad, N. (2015) ‘A Fluidic Colorimetric Sensor Design for Water Hardness Detection’, *Ieee Sensors Journal*, 15(2), pp. 819–826. doi: 10.1109/Jsen.2014.2351813.
- Bishop, C. M. (2006) *Pattern recognition and machine learning*. Springer.
- Brownlee, J. (2018) *Better Deep Learning: Train Faster, Reduce Overfitting, and Make Better*. v1.5. Machine Learning Mastery. Available at: https://books.google.co.uk/books/about/Better_Deep_Learning.html?id=T1-nDwAAQBAJ&redir_esc=y (Accessed: 30 January 2020).
- Bu, S. *et al.* (2018) ‘A pregnancy test strip for detection of pathogenic bacteria by using concanavalin A-human chorionic gonadotropin-Cu₃(PO₄)₂ hybrid nanoflowers, magnetic separation, and smartphone readout’, *Microchimica Acta*. Springer Vienna, 185(10), p. 464. doi: 10.1007/s00604-

018-2968-2.

Chang, C.-C. *et al.* (2019) 'Gold Nanoparticle-Based Colorimetric Strategies for Chemical and Biological Sensing Applications', *Nanomaterials*. MDPI AG, 9(6), p. 861. doi: 10.3390/nano9060861.

Chen, W. *et al.* (2016) 'A Simple Paper-Based Colorimetric Device for Rapid Mercury(II) Assay', *Scientific Reports*. Nature Publishing Group, 6(August). doi: 10.1038/srep31948.

Choodum, A., Sriprom, W. and Wongniramaikul, W. (2019) 'Portable and selective colorimetric film and digital image colorimetry for detection of iron', *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. Elsevier, 208, pp. 40–47. doi: 10.1016/J.SAA.2018.09.062.

Cooper, D. *et al.* (2012) 'Mobile Image Ratiometry: A New Method for Instantaneous Analysis of Rapid Test Strips', in *Nature Precedings*, pp. 2–3. doi: 10.1038/npre.2012.6827.1.

Coskun, Ahmet F. Wong, J. *et al.* (2012) 'A personalized food allergen testing platform on a cellphone', 100(2), pp. 130–134. doi: 10.1016/j.pestbp.2011.02.012. Investigations.

Enzo Life Sciences inc. (2015) *Enzo ELISA Plate Reader, Android App on Google Play*. Available at: <https://play.google.com/store/apps/details?id=com.enzo.elisaplatereader> (Accessed: 21 September 2017).

Facchini, F. A. *et al.* (2018) 'Structure–Activity Relationship in Monosaccharide-Based Toll-Like Receptor 4 (TLR4) Antagonists', *Journal of Medicinal Chemistry*. American Chemical Society, 61(7), pp. 2895–2909. doi: 10.1021/acs.jmedchem.7b01803.

Feng, S. *et al.* (2014) 'Immunochromatographic diagnostic test analysis using google glass', *ACS Nano*, pp. 3069–3079.

Guyon, I., Elisseeff, A. and De, A. M. (2003) 'An Introduction to Variable and Feature Selection', *Journal of Machine Learning Research*, 3, pp. 1157–1182. Available at: <http://jmlr.csail.mit.edu/papers/volume3/guyon03a/guyon03a.pdf> (Accessed: 22 May 2018).

He, K. *et al.* (2016) 'Deep residual learning for image recognition', in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, pp. 770–778. doi: 10.1109/CVPR.2016.90.

Hoque Tania, M. (2018) *An Intelligent Image-based Colourimetric Test Framework for Diagnosis*. Anglia Ruskin University. Available at: <http://arro.anglia.ac.uk/id/eprint/704504> (Accessed: 3 December 2019).

Hoque Tania, M. *et al.* (2020) 'Intelligent image-based colourimetric tests using machine learning framework for lateral flow assays', *Expert Systems with Applications*. Pergamon, 139, p. 112843. doi: 10.1016/J.ESWA.2019.112843.

Huang, G. *et al.* (2016) 'Deep networks with stochastic depth', in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Verlag, pp. 646–661. doi: 10.1007/978-3-319-46493-0_39.

ImageNet (2016) *ImageNet*. Available at: <http://image-net.org/index> (Accessed: 28 July 2018).

Karisen, H. and Dong, T. (2016) 'Illumination and device independence for colorimetric detection of urinary biomarkers with smartphone', in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, pp. 5184–5187. doi: 10.1109/EMBC.2016.7591895.

Karlsen, H. (2018) *Smartphone-based urinary biomarker detection: an application-oriented device and algorithm*. University College of Southeast Norway. Available at: <https://brage.bibsys.no/xmlui/handle/11250/2480144> (Accessed: 21 March 2018).

- Kim, H. *et al.* (2017) 'Colorimetric analysis of saliva--alcohol test strips by smartphone-based instruments using machine-learning algorithms', *Appl. Opt.*, 56(1), pp. 84–92. doi: 10.1364/AO.56.000084.
- Konnaiyan, K. *et al.* (2017) 'mHealth Dipstick Analyzer For Monitoring of Pregnancy Complications', *IEEE Sensors Journal*, 17(22), pp. 7311–7316.
- Koo, K. M., Wee, E. J. H. and Trau, M. (2016) 'Colorimetric *TMPRSS2-ERG* Gene Fusion Detection in Prostate Cancer Urinary Samples via Recombinase Polymerase Amplification', *Theranostics*, 6(9), pp. 1415–1424. doi: 10.7150/thno.15250.
- Koydemir, H. C. *et al.* (2015) 'Rapid imaging, detection and quantification of *Giardia lamblia* cysts using mobile-phone based fluorescent microscopy and machine learning.', *Lab on a chip*, 15(5), pp. 1284–93. doi: 10.1039/c4lc01358a.
- Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012) 'ImageNet Classification with Deep Convolutional Neural Networks', in *Advances in neural information processing systems*. Neural Information Processing Systems Foundation, Inc., pp. 1097–1105. Available at: <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf> (Accessed: 9 April 2018).
- de la Rica, R. and Stevens, M. M. (2012) 'Plasmonic ELISA for the ultrasensitive detection of disease biomarkers with the naked eye', *Nature Nanotechnology*. Nature Publishing Group, 8(9), pp. 1759–1764. doi: 10.1038/nnano.2012.186.
- Lopez-Ruiz, N. *et al.* (2014) 'Smartphone-based simultaneous pH and nitrite colorimetric determination for paper microfluidic devices', *Analytical Chemistry*, 86(19), pp. 9554–9562. doi: 10.1021/ac5019205.
- Magiati, M. *et al.* (2019) 'Lateral flow test for meat authentication with visual detection', *Food Chemistry*. Elsevier, 274, pp. 803–807. doi: 10.1016/J.FOODCHEM.2018.09.063.
- Mane, Viraj; Yakub, S. I. (2018) 'Detection of flu using thermal imaging - Google Patents'. US. Available at: <https://patents.google.com/patent/US20190192010A1/en> (Accessed: 29 January 2020).
- Marcelino, P. (2018) *Transfer learning from pre-trained models, Towards Data Science*. Available at: <https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751> (Accessed: 1 May 2020).
- Mathisen, B. M. *et al.* (2019) 'Learning similarity measures from data', *Progress in Artificial Intelligence*. Springer Verlag, pp. 1–15. doi: 10.1007/s13748-019-00201-2.
- MATLAB & Simulink (2018a) *Deep Dream Images Using AlexNet, MathWorks*. Available at: <https://uk.mathworks.com/help/nnet/examples/deep-dream-images-using-alexnet.html> (Accessed: 9 September 2018).
- MATLAB & Simulink (2018b) *Preprocess Images for Deep Learning, MathWorks*. Available at: <https://uk.mathworks.com/help/deeplearning/ug/preprocess-images-for-deep-learning.html;jsessionid=e43272072335d60313e2016301a5> (Accessed: 15 September 2018).
- Matthews, J. *et al.* (2012) 'Rapid dengue and outbreak detection with mobile systems and social networks', *Mobile Networks and Applications*, 17(2), pp. 178–191. doi: 10.1007/s11036-011-0295-5.
- Mordvintsev, Alexander; Olah, Christopher; Tyka, M. (2015) *Inceptionism: Going Deeper into Neural Networks, Google*. Available at: <https://web.archive.org/web/20150703064823/http://googleresearch.blogspot.co.uk/2015/06/inceptionism-going-deeper-into-neural.html> (Accessed: 9 September 2018).
- Mordvintsev, Alexander; Olah, C. and Tyka, M. (2015) 'DeepDream'. Google. Available at: <https://web.archive.org/web/20150708233542/http://googleresearch.blogspot.co.uk/2015/07/deepdrea>

m-code-example-for-visualizing.html.

Murphy, K. P. (2012) *Machine learning : A probabilistic perspective*. MIT Press. Available at: <https://mitpress.mit.edu/books/machine-learning-1> (Accessed: 29 July 2018).

Mutlu, A. Y. *et al.* (2017) 'Smartphone-based colorimetric detection via machine learning', *The Analyst*, 142(13), pp. 2434–2441. doi: 10.1039/C7AN00741H.

Nair, V. and Hinton, G. E. (2010) 'Rectified linear units improve restricted boltzmann machines', in *27th International Conference on International Conference on Machine Learning*. Haifa, Israel: Association for Computing Machinery, pp. 807–814. Available at: <https://dl.acm.org/citation.cfm?id=3104425> (Accessed: 16 September 2018).

Nash, W., Drummond, T. and Birbilis, N. (2018) 'A review of deep learning in the study of materials degradation', 2, p. 37. doi: 10.1038/s41529-018-0058-x.

Ng, A. (2018) *deeplearning.ai, coursera*. Available at: <https://www.deeplearning.ai/> (Accessed: 15 September 2018).

Pena-Pereira, F. *et al.* (2020) 'A paper-based colorimetric assay with non-instrumental detection for determination of boron in water samples', *Talanta*. Elsevier B.V., 208, p. 120365. doi: 10.1016/j.talanta.2019.120365.

Promphet, N. *et al.* (2019) 'Non-invasive textile based colorimetric sensor for the simultaneous detection of sweat pH and lactate', *Talanta*. Elsevier, 192, pp. 424–430. doi: 10.1016/J.TALANTA.2018.09.086.

Qin, X. *et al.* (2015) 'A colorimetric chemical sensing platform for real-time monitoring of indoor formaldehyde', *IEEE Sensors Journal*, 15(3), pp. 1545–1551. doi: 10.1109/JSEN.2014.2364142.

Rahmat, R. F. *et al.* (2018) 'Automated color classification of urine dipstick image in urine examination', *Journal of Physics: Conference Series*. IOP Publishing, 978(1), p. 012008. doi: 10.1088/1742-6596/978/1/012008.

Sahlsten, J. *et al.* (2019) 'Deep Learning Fundus Image Analysis for Diabetic Retinopathy and Macular Edema Grading', *Scientific Reports*. Nature Publishing Group, 9(1). doi: 10.1038/s41598-019-47181-w.

Santos Neto, J. H. *et al.* (2018) 'Speciation analysis based on digital image colorimetry: Iron (II/III) in white wine', *Talanta*. Elsevier. doi: 10.1016/J.TALANTA.2018.09.102.

Scherer, D., Müller, A. and Behnke, S. (2010) 'Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition', in *20th International Conference on Artificial Neural Networks (ICANN)*. Thessaloniki, Greece. Available at: <http://www.ais.uni-bonn.de> (Accessed: 14 September 2018).

Serrano, J. M. *et al.* (2016) 'Colorimetric Analysis for On-Line Arc-Welding Diagnostics by Means of Plasma Optical Spectroscopy', *IEEE Sensors Journal*, 16(10), pp. 3465–3471. doi: 10.1109/JSEN.2015.2413057.

Shabut, A. M. *et al.* (2018) 'An intelligent mobile-enabled expert system for tuberculosis disease diagnosis in real time', *Expert Systems with Applications*. Pergamon, 114, pp. 65–77. doi: 10.1016/j.eswa.2018.07.014.

Shin, J. *et al.* (2017) 'Smart Forensic Phone: Colorimetric analysis of a bloodstain for age estimation using a smartphone', *Sensors and Actuators B: Chemical*. Elsevier B.V., 243, pp. 221–225. doi: 10.1016/j.snb.2016.11.142.

Sicard, C. *et al.* (2015) 'Tools for water quality monitoring and mapping using paper-based sensors and cell phones', *Water Research*, 70, pp. 360–369. doi: 10.1016/j.watres.2014.12.005.

- Sicasys Software GmbH (2017) *Spotxel® Reader, Google Play*. Available at: <https://play.google.com/store/apps/details?id=com.sicasys.spotxel&hl=en> (Accessed: 12 January 2018).
- Smith, G. T. *et al.* (2016) 'Robust dipstick urinalysis using a low-cost, micro-volume slipping manifold and mobile phone platform', *Lab Chip*. Royal Society of Chemistry, 58, pp. 951–954. doi: 10.1039/C6LC00340K.
- Smith, J. E. *et al.* (2014) 'Colorimetric detection with aptamer-gold nanoparticle conjugates coupled to an android-based color analysis application for use in the field', *Talanta*, 121, pp. 247–55. doi: 10.1016/j.talanta.2013.12.062.
- Solmaz, M. E. *et al.* (2018) 'Quantifying colorimetric tests using a smartphone app based on machine learning classifiers', *Sensors and Actuators B: Chemical*. Elsevier, 255, pp. 1967–1973. doi: 10.1016/J.SNB.2017.08.220.
- Szegedy, C. *et al.* (2015) 'Going deeper with convolutions', in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Boston, MA: IEEE Computer Society, pp. 1–9. doi: 10.1109/CVPR.2015.7298594.
- Szegedy, C. *et al.* (2016) 'Rethinking the Inception Architecture for Computer Vision', in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. doi: 10.1109/CVPR.2016.308.
- Tania, M. H. *et al.* (2017) 'An Automated Colourimetric Test by Computational Chromaticity Analysis: A Case Study of Tuberculosis Test', in *Advances in Intelligent Systems and Computing*. Springer, Cham, pp. 313–320. doi: 10.1007/978-3-319-60816-7.
- Tania, M. H. *et al.* (2018) 'Clustering and Classification of a Qualitative Colorimetric Test', in *2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*. Southend, United Kingdom: IEEE, pp. 7–11. doi: 10.1109/iCCECOME.2018.8658480.
- Tiulpin, A. *et al.* (2019) 'Multimodal Machine Learning-based Knee Osteoarthritis Progression Prediction from Plain Radiographs and Clinical Data'. doi: 10.1038/s41598-019-56527-3.
- Vashist, S. K. *et al.* (2015) 'A smartphone-based colorimetric reader for bioanalytical applications using the screen-based bottom illumination provided by gadgets', *Biosensors and Bioelectronics*. Elsevier, 67, pp. 248–255. doi: 10.1016/j.bios.2014.08.027.
- Wang, S. *et al.* (2018) 'Comprehensive analysis of lung cancer pathology images to discover tumor shape and boundary features that predict survival outcome', *Scientific Reports*. Nature Publishing Group, 8(1). doi: 10.1038/s41598-018-27707-4.
- Wang, T. T. *et al.* (2020) 'A feasible image-based colorimetric assay using a smartphone RGB camera for point-of-care monitoring of diabetes', *Talanta*. Elsevier B.V., 206, p. 120211. doi: 10.1016/j.talanta.2019.120211.
- Wang, Y. *et al.* (2016) 'A smartphone-based colorimetric reader coupled with a remote server for rapid on-site catechols analysis', *Talanta*. Elsevier, 160, pp. 194–204. doi: 10.1016/j.talanta.2016.07.012.
- Wei, Q. *et al.* (2014) 'Detection and spatial mapping of mercury contamination in water samples using a smart-phone', *ACS Nano*, 8(2), pp. 1121–1129. doi: 10.1021/nn406571t.
- Welch, H. G. and Albertsen, P. C. (2009) 'Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986–2005', *Journal of the National Cancer Institute*, 101(19), pp. 1325–1329. doi: 10.1093/jnci/djp278.
- Wen, S. H. *et al.* (2019) 'Colorimetric Assay Conversion to Highly Sensitive Electrochemical Assay for Bimodal Detection of Arsenate Based on Cobalt Oxyhydroxide Nanozyme via Arsenate

Absorption', *Analytical chemistry*. NLM (Medline), 91(10), pp. 6487–6497. doi: 10.1021/acs.analchem.8b05121.

Wongsrichanalai, C. *et al.* (2007) 'A review of malaria diagnostic tools: Microscopy and rapid diagnostic test (RDT)', *American Journal of Tropical Medicine and Hygiene*, 77(SUPPL. 6), pp. 119–127. doi: 10.3126/ajms.v1i2.2965.

World Health Organization (2009) *Screening assays - Screening Donated Blood for Transfusion-Transmissible Infections: Recommendations*. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK142997/> (Accessed: 5 May 2020).

Yan, S. *et al.* (2018) 'Identification of aminoglycoside antibiotics in milk matrix with a colorimetric sensor array and pattern recognition methods', *Analytica Chimica Acta*. Elsevier. doi: 10.1016/J.ACA.2018.06.004.

Yetisen, A. K. *et al.* (2014) 'A smartphone algorithm with inter-phone repeatability for the analysis of colorimetric tests', *Sensors and Actuators B: Chemical*. Elsevier B.V., 196, pp. 156–160. doi: 10.1016/j.snb.2014.01.077.

Yu, L. *et al.* (2015) 'Disposable lateral flow-through strip for smartphone-camera to quantitatively detect alkaline phosphatase activity in milk', *Biosensors and Bioelectronics*. Elsevier, 69, pp. 307–315. doi: 10.1016/j.bios.2015.02.035.

Zheng, L. *et al.* (2018) 'A microfluidic colorimetric biosensor for rapid detection of Escherichia coli O157:H7 using gold nanoparticle aggregation and smart phone imaging', *Biosensors and Bioelectronics*. Elsevier. doi: 10.1016/J.BIOS.2018.10.006.